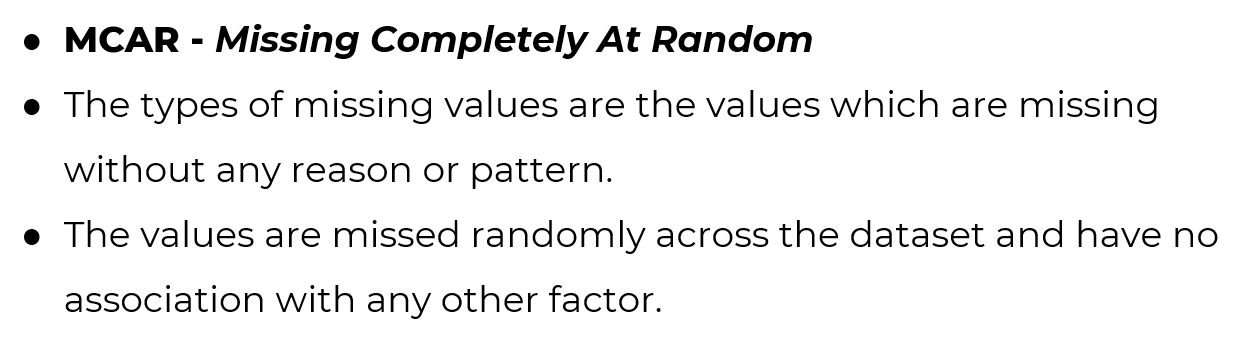
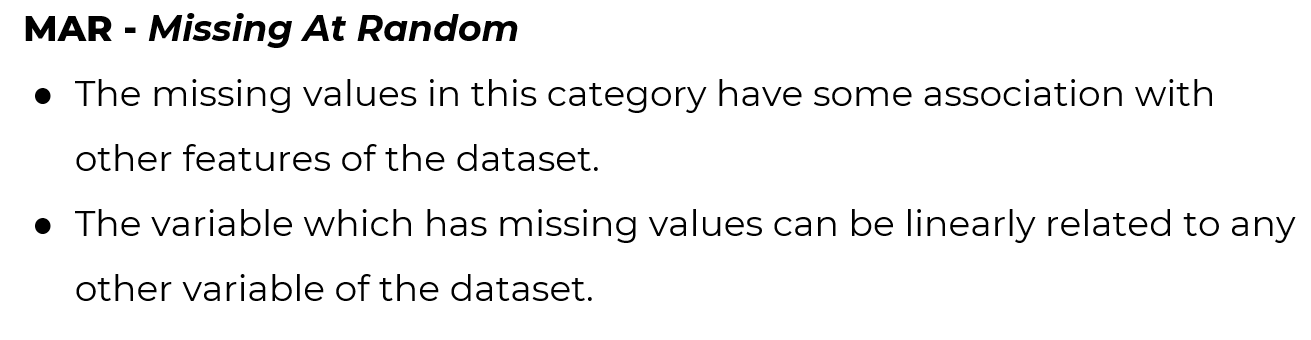
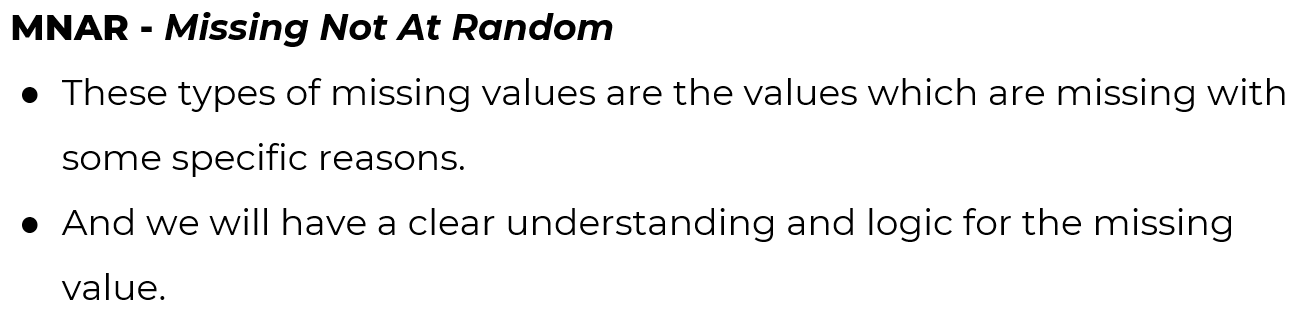
* There are **3 types** of missing values:







* When a column or row missing value % is very high, we can **delete** that. The % depends on dataset size. If smaller dataset we can set it to 20%, in case of large dataset we can set it to 40%. Note: these rules can change in different scenario. For example,
  1. If the target has high correlation (relation between two variable) with the column.
  2. If the removal can lead to information loss.
* **Imputing(representing/replacing) missing values using business logic**:

Business logic is simply relying on business implication or business understanding of solving a problem. For example, if a customer product review rating is missing, we can’t just put random number there or average rating. This would harm the dataset. In this case, we can assign -1 or -999 to the data.

* **Imputing values using MeanMedianMode:** 
  1. **Mode function:** Returns most frequently occurring data.
  2. **Median function:** Median can be defined as the middle number of a group of numbers. That is, half the numbers are greater than the median, and half the numbers are less.

We should use it when there are **outliers** present in data. Median function isn’t sensitive towards outliers.

Note: outliers = data point that differ significantly from other measurement

* 1. **Mean function:** We should use it when there is no outlier in the data, as this function is very sensitive towards outlier.